PB86-162435

Pattern Recognition Analysis of a Set of Mutagenic Aliphatic N-Nitrosamines

(U.S.) Health Effects Research Lab.
Research Triangle Park, NC

1985

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA (Please read Instructions on the reverse before completir)		PB86-162435
1. REPORT NO. EPA/600/J-85/299	2.	3. F
4. TITLE AND SUBTITLE Pattern Recognition Analysis of a Set of Mutagenic Aliphatic N-Nitrosamines (Journal Version)		5. REPORT DATE 1985 6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) S. Mesnow, R. Langenbach, M. Mass		8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Health Effects Research Laboratory USEPA Research Triangle Park, NC 27711		10. PROGRAM ELEMENT NO. 11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Office of Research & Development Health Effects Research Laboratory USEPA Research Triangle Park, NC 27711		13. TYPE OF REPORT AND PERIOD COVERED 14. SPONSORING AGENCY CODE EPA-600/11
15. SUPPLEMENTARY NOTES Environmental Health Perspectives Vol. 61, pp. 345-349, 1985 (September)		

16. ABSTRACT

A set of 21 mutagenic aliphatic N-nitrosamines were subjected to a pattern recognition analysis using ADAPT software. Four descriptors based on molecular connectivity, geometry and sigma charge on nitrogen were capable of achieving a 100% classification using the linear learning machine or iterative least squares algorithms. Three descriptors were capable of a 90.5% and two descriptors of a 85.7% overall correct classification. Three of the four descriptors were each capable of classifying 15 of the 16 active chemicals while it required three of the four descriptors to correctly classify two of the five inactive chemicals. These results are in concert with previous observations that molecular connectivity, geometry, and sigma charge on nitrogen are powerful descriptors for separating active from inactive mutagenic and carcinogenic N-nitrosamines.

17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
RELEASE TO PUBLIC	UNCLASSIFIED	6
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

EPA/600/J-85/299
JOURNAL ARTICLE
SEPT
Environmental Health Perspectives
Vol. 61, pp. 345-349, 1985

Pattern Recognition Analysis of a Set of Mutagenic Aliphatic *N*-Nitrosamines

by Stephen Nesnow,*† Robert Langenbach,‡ and Marc J. Mass*

A set of 21 mutagenic aliphatic *N*-nitrosamines were subjected to a pattern recognition analysis using ADAPT software. Four descriptors based on molecular connectivity, geometry and sigma charge on nitrogen were capable of achieving a 100% classification using the linear learning machine or iterative least squares algorithms. Three descriptors were capable of a 90.5% and two descriptors of a 85.7% overall correct classification. Three of the four descriptors were each capable of classifying 15 of the 16 active chemicals while it required three of the four descriptors to classify correctly two of the five inactive chemicals. These results are in concert with previous observations that molecular connectivity, geometry, and sigma charge on nitrogen are powerful descriptors for separating active from inactive mutagenic and carcinogenic *N*-nitrosamines.

Introduction

Structure–activity analysis is a powerful tool to study the relationship between chemicals and their biological effects. The elucidation of specific chemical features that directly or indirectly result in the expression of a biochemical or biological change can provide new insights into the mechanisms of action of the chemicals as well as the structure and function of the biological system.

We have previously reported on the development of an innovative genetic toxicology system that determines the mutagenic activity of chemicals in mammalian cells using a specific gene mutation at the sodium/potassium–adenosine triphosphatase (ATPase) gene (*1*). The system utilizes mutable Chinese hamster lung cells (V79) that are cocultivated with primary hamster hepatocytes serving as the metabolic activation component. The use of intact liver cells rather than subcellular liver enzyme fractions allows the simulation of the metabolic competancy of liver *in vivo* and as a result quantitative relationships observed *in vivo* can be reproduced *in vitro* (*2,3*). Intact liver cells also provide a means to study liver specific chemical toxicants (*4,5*). In this regard, aliphatic *N*-nitrosamines have been recognized as hepatotoxic and hepatocarcinogenic in a variety of rodent species including mouse, rat, hamster, and guinea pig (*6*). The mechanisms by which some of these *N*-nitrosamines (e.g., dimethylnitrosamine, diethylnitrosamine) are metabolically activated to genotoxic and carcinogenic forms are well known and involve α-carbon oxidation and the formation of a methyl or ethyl carbonium ion which alkylates DNA. However, the mechanism by which longer chain aliphatic *N*-nitrosamines are metabolically activated and the species that interact with DNA are still a matter of speculation. It has been postulated that propyl *N*-nitrosamines are metabolized to both a methylating and propylating species as methyl residues in DNA (*7*) and propyl residues in RNA have been isolated (*8*). Dipropylnitrosamine is also metabolized to bis-2-oxypropylnitrosamine (*7–9*), which can methylate DNA *in vivo* after further metabolism (*10,11*).

Previous structure–activity studies on groups of carcinogenic *N*-nitrosamines by Rose and Jurs (*12*) using pattern recognition and by Dunn and Wold (*13,14*) using SIMCA pattern recognition have indicated a variety of functions which participate in the separation of positive and negative chemicals. In this study a set of related propyl and butyl *N*-nitrosamines previously examined for their ability to induce specific gene mutation in the V79 cell-hamster hepatocyte cocultivation system (*11*) were incorporated into a structure activity analysis using pattern recognition techniques. It was of interest to relate specific physiochemical, electronic, and topo-

*Carcinogenesis and Metabolism Branch, U.S. Environmental Protection Agency, Health Effects Research Laboratory, Research Triangle Park, NC 27711.
‡National Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709.
†To whom requests for reprints should be addressed.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

graphic parameters to mutagenic activity in order to study the mechanisms of action of these *N*-nitrosamines.

Methods

Computational Methods

The methods used were implemented through an interactive computer software system known as ADAPT (Automated Data Analysis by Pattern-recognition Techniques), which has been reported elsewhere (*15*). ADAPT is maintained on a UNIVAC 1100 mainframe and interactively accessed via a Tektronix 4054 terminal using PLOT 10 graphics.

The chemical structures of the compounds comprising the data set were entered into computer disc files by sketching the structures on a graphics display terminal using a joystick. Each structure was modeled using a force-field molecular mechanics model builder.

Descriptors were calculated using various algorithms incorporated into the ADAPT system. Molecular connectivity was calculated according to Kier and Hall (*16*), a sigma charge according to del-Ré (*17*), and geometric descriptors from atomic coordinates.

Data Set

The 21 *N*-nitrosamines (Fig. 1) were evaluated for their ability to mutate Chinese hamster lung cells (V79) in the presence of primary hamster hepatocytes from Syrian golden hamsters according to previously published procedures (*1,2,11*). The mutation locus selected was ouabain resistance. The data for 14 of the 21 *N*-nitrosamines has already been reported (*11*). MTBN, 2BOB, 2BHB, DEN, 3POP, 3MHP and 3MOP were evaluated under the same conditions reported by Langenbach et al. (*1*), and their activity (active or inactive) ascribed from the mutation frequency [No. of ouabain resistant mutants/10^6 survivors] at 0.7 mM (< 10, inactive; > 10 active).

Results and Discussion

A set of 21 *N*-nitrosamines consisting of 14 propylnitrosamines, 5 butylnitrosamines, dimethylnitrosamine, and diethylnitrosamine evaluated for mutagenic activity in a hepatocyte-V79 cell mutagenesis assay were used (Table 1, Fig. 1).

Of the 21 *N*-nitrosamines, 16 had mutagenic activity when tested at a concentration of 0.7 mM. Mutagenic activity was defined as the induction of more than 10 ouabain-resistant mutants per 10^6 surviving V79 cells. Five *N*-nitrosamines (3HPP, BAP, BHP, MTBN, and 2BHB) produced less than the above number of mutants and were considered inactive.

The molecular structures of the *N*-nitrosamine dataset were entered into the ADAPT pattern recognition system and descriptors were generated. Initially a large

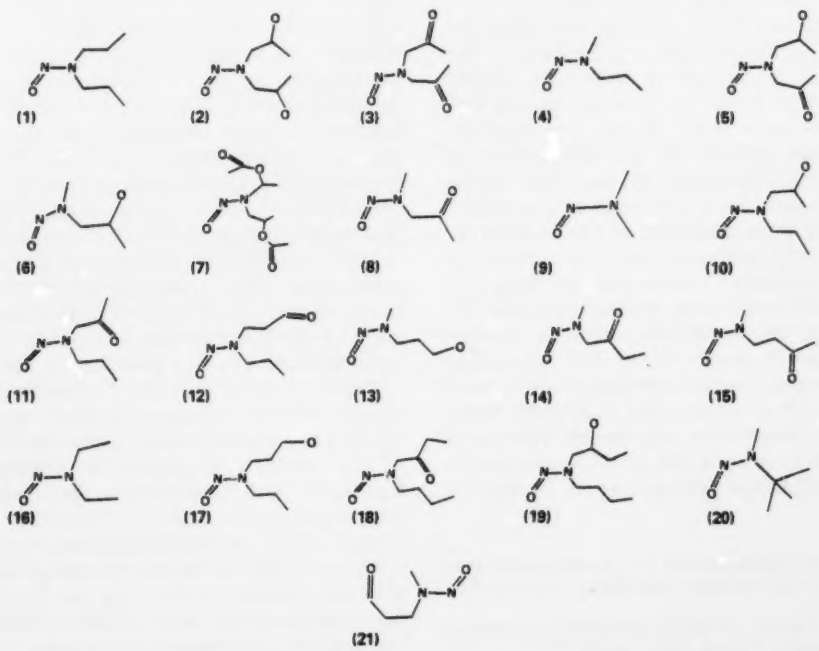

FIGURE 1. Structures of 21 *N*-nitrosamines used in pattern recognition analysis. The numbers refer to the index numbers in Table 1.

Table 1. *N*-Nitrosamines used in pattern recognition analyses.

Index	Nitrosamine	Abbreviation	Mutagenicity[a]
1	Dipropylnitrosamine	DPN	+
2	Bis-2-hydroxypropylnitrosamine	BHP	-
3	Bis-2-oxypropylnitrosamine	BOP	+
4	Methylpropylnitrosamine	MPN	+
5	2-Hydroxypropyl-2-oxypropylnitrosamine	HPOP	+
6	Methyl-2-hydroxypropylnitrosamine	MHP	+
7	Bis-2-acetoxypropylnitrosamine	BAP	-
8	Methyl-2-oxypropylnitrosamine	MOP	+
9	Dimethylnitrosamine	DMN	+
10	2-Hydroxypropylpropylnitrosamine	HPP	+
11	Propyl-2-oxypropylnitrosamine	POP	+
12	Propyl-3-oxypropylnitrosamine	3POP	+
13	Methyl-3-hydroxypropylnitrosamine	3MHP	+
14	Methyl-2-oxybutylnitrosamine	2MOP	+
15	Methyl-3-oxybutylnitrosamine	3MOB	+
16	Diethylnitrosamine	DEN	+
17	3-Hydroxypropylpropylnitrosamine	3HPP	-
18	Butyl-2-oxybutylnitrosamine	2BOB	+
19	Butyl-2-hydroxybutylnitrosamine	2BHB	-
20	Methyl-*t*-butylnitrosamine	MTBN	-
21	Methyl-3-oxypropylnitrosamine	3MOP	+

[a]Mutagenicity determined in the hamster hepatocyte-V79 cell cocultivation system using ouabain resistance as the genetic marker.

Table 2. Descriptors used for pattern recognition analyses.[a]

Descriptor	Overall	Active class	Inactive class
MOLC1[b]	3.88 ± 1.05	$3.57 \pm 7.48 \times 10^{-1}$	4.88 ± 1.33
GEOM3[c]	$9.32 \times 10^{-2} \pm 9.97 \times 10^{-2}$	$6.95 \times 10^{-2} \pm 6.03 \times 10^{-2}$	$1.69 \times 10^{-1} \pm 1.63 \times 10^{-1}$
GEOM6[d]	$2.29 \times 10^{1} \pm 1.31 \times 10^{1}$	$2.11 \times 10^{1} \pm 8.44$	$2.86 \times 10^{1} \pm 2.31 \times 10^{1}$
SCSA1[e]	$-1.438 \times 10^{-1} \pm 1.663 \times 10^{-3}$	$-1.437 \times 10^{-1} \pm 1.670 \times 10^{-3}$	$-1.442 \times 10^{-1} \pm 1.753 \times 10^{-3}$

[a]Values are mean ± SD.
[b]MOLC1 is path 1 molecular connectivity for all bonds.
[c]GEOM3 is the smallest principal moment (Z).
[d]GEOM6 is the ratio of the intermediate principal moment (Y) and the smallest principal moment (Z).
[e]SCSA1 is the sigma charge on the terminal nitrogen to which the nitroso moiety is attached (*N-N=O).

number of descriptors were calculated, including molecular fragment descriptors (number and kinds of atoms and bond types), branching index descriptors, molecular size and volumes descriptors, and charge distribution descriptors. In order to insure that each descriptor contributed unique information about the molecules, multiple correlation analyses of the descriptors were performed. Those descriptors whose values were highly correlated with other descriptors were discarded.

Four descriptors (Table 2) were found to be free of colinearity with the most correlated descriptors GEOM3 and GEOM6 exhibiting a correlation coefficient of 0.576 (Table 3). No multicolinearity was detected. MOLC1 is the path 1 molecular connectivity which is an index of branching and a measure of the complexity of the molecule; the geometric descriptor GEOM3 describes the smallest of the three principal moments of a molecule; GEOM6 is the ratio between the intermediate and smallest principal moments; SCSA1 was calculated by the del-Ré method and is the sigma charge on the terminal nitrogen to which the nitroso moiety is attached

Table 3. Noncolinearity of descriptors.[a]

	GEOM3	GEOM6	SCSA1
MOLC1[b]	0.325	0.289	0.256
GEOM3	—	0.576	0.316
GEOM6	—	—	0.038

[a]Values are the correlation coefficient (R).
[b]Refer to Table 2 and the Results section for a explanation of descriptors.

Table 4. Descriptor contribution by specific exclusion analysis.[a]

Descriptor left out	Active chemicals misclassified	Inactive chemicals misclassified	% Correct[b]
MOLC1[c]	HPOP	3HPP	90.5
GEOM3	DPN	MTBN, BHP	85.7
GEOM6	BOP	MTBN, BHP	85.7
SCSA1	HPOP, BOP	MTBN	85.7

[a]Calculated by using the 21 *N*-nitrosamine dataset with the Linear Learning Machine routine.
[b]Weighted average of percent correctly identified active and inactive chemicals.
[c]Refer to Table 2 and Results section for an explanation of descriptors.

Table 5. Classification results with some pattern-recognition routines.

Routine	% Correct classification		
	Active	Inactive	Overall
Linear learning machine	100 (16/16)[a]	100 (5/5)[b]	100[c]
Bayes linear classifier	81.3 (13/16)	80 (4/5)	81
Bayes quadratic classifier	93.8 (15/16)	80 (4/5)	90.5
Iterative least squares	100 (16/16)	100 (5/5)	100
Linear discriminant analysis	100 (16/16)	80 (4/5)	95.2

[a] Number correctly classified active chemicals/number in active class.
[b] Number correctly classified inactive chemicals/number in inactive class.
[c] Weighted average of percent correctly identified active and inactive chemicals.

(*N-N = O). The simple statistics of the distributions of the four descriptors partitioned into the active and inactive classes are given in Table 2.

Using the linear learning machine algorithm, the entire dataset and four descriptors were first used to train weight vectors for construction of a separable pattern space. Complete convergence was attained rapidly, usually with less than 50 feedbacks using a deadzone value of 100. Using 100 training sets of randomly generated datasets where three of the active class and one of the inactive class were removed, the linear learning machine algorithm achieved 97% overall correct prediction of 17 member prediction datasets. When the prediction datasets consisted of the four members left out of the training dataset, the overall prediction was 80%.

The database of 21 *N*-nitrosamines did not contain an equal number of active and inactive chemicals, and the number of descriptors used to achieve separation was only one less than the number of inactive chemicals. It was necessary to determine if the four descriptors chosen specifically classified only the inactive compounds or if they recognized both the active and inactive compounds. This possibility was assessed by excluding each of the descriptors separately to determine the effectiveness of the linear learning machine. Exclusion of any of the four descriptors resulted in nonconvergence of the training routine of the linear learning machine; the weight vectors generated could not separate the active and inactive compounds during training without misclassification. However, the exclusion of no one descriptor specifically resulted in misclassification of only an inactive compound (Table 4). Omission of a descriptor always resulted in an error in classifying at least one active and one inactive *N*-nitrosamine. This would suggest that although the number of descriptors used was nearly equal to the number of inactive *N*-nitrosamines, each descriptor did not select for a particular member of the inactive class. Indeed, it was found instead that in order to correctly classify the inactive *N*-nitrosamines MTBN and BHP, a minimum of three and two descriptors were necessary, respectively. The molecular connectivity and geometric descriptors were capable of correctly classifying 15 of 16 active chemicals, indicating each to strongly contribute to the separation. HPOP and BOP were the most difficult to identify as it required three descriptors to correctly identify them (Table 4).

Utilizing four descriptors and the complete 21 chemical dataset the linear learning machine achieved an overall classification of 100% (Table 5). The iterative least-squares algorithm was also able to attain 100% correct overall classification in 71 iterations. A linear discriminant function analysis program was able to generate a discriminant function that could separate the active class correctly but misclassified one chemical in the inactive class. Bayes' quadratic classifier achieved a 93.8% correct classification of the active class and incorrectly classified an inactive class member. Bayes' linear classifier was least effective at separating either the inactive or active class and classified 13/16 (81.2%) of the active class correctly but missed 1/5 of the inactive class.

The power of the descriptors chosen is evident when only two are used in the linear learning machine classification of the 21 *N*-nitrosamine dataset. The most accurate two descriptor set was MOLC1/GEOM3 which gave an overall classification of 85.7%. MOLC1/GEOM6 or GEOM3/GEOM6 gave identical classifications, 80.9%. All other combinations produced classifications of < 71%.

Earlier structure–activity analyses using 14 of these 21 *N*-nitrosamines indicated that molecular connectivity path 1 calculations correlated well with the absolute mutation frequencies of the 14 *N*-nitrosamines (*11*). Rose and Jurs (*12*) also found molecular connectivity to play an important role in the classification of a 150 *N*-nitrosamine carcinogen dataset as its deletion from the descriptor list reduced the overall classification from 97.33% to 74.67%. Rose and Jurs (*12*) also found GEOM3 (smallest principal moment) and sigma charge on the *N*-nitrosamine moiety to be important in their pattern recognition analyses. Therefore, the molecular connectivity, sigma charge, and geometric descriptors seem to be a powerful combination for separating active from inactive mutagenic and carcinogenic *N*-nitrosamines.

The authors thank Carol Evans for her assistance with the computer programming and ADP support, Joye Denning for preparation of this manuscript, and Peter Jurs, Pennsylvania State University for the ADAPT software.

REFERENCES

1. Langenbach, R., Freed, H. J., and Huberman, E. Liver cell-mediated mutagenesis of mammalian cells by liver carcinogens. Proc. Natl. Acad. Sci. (U.S.) 75: 2864–2867 (1978).
2. Langenbach, R., Gingell, R., Kuszynski, C., Walker, B., Nagel, D., and Pour, P. Mutagenic activities of oxidized derivatives of *N*-nitrosodipropylamine in the liver cell-mediated and *Salmonella typhimurium* assays. Cancer Res. 40: 3463–3467 (1980).
3. Jones, C. A., Marlino, P. J., Lijinsky, W., and Huberman, E. The relationship between the carcinogenicity and mutagenicity of nitrosamines in a hepatocyte-mediated mutagenicity assay. Carcinogenesis 2: 1075–1077 (1981).
4. Langenbach, R., Nesnow, S., Tompa, A., Gingell, R., and Kuszynski, C. Lung and liver cell-mediated mutagenesis systems: specificities in the activation of chemical carcinogens. Carcinogenesis 2: 851–858 (1981).
5. Langenbach, R., and Nesnow, S. Cell-mediated mutagenesis, an approach to studying organ specificity of chemical carcinogens. In: Organ and Species Specificity in Chemical Carcinogenesis (R. Langenbach, S. Nesnow and J. Rice, Eds.), Plenum Press, New York, 1983, pp. 377–390.
6. Magee, P. N., Montesano, R., and Preussmann, R. *N*-Nitroso compounds and related carcinogens. In: Chemical Carcinogens, ACS Monograph No. 173 (C. E. Searle, Ed.), American Chemical Society, Washington, DC, 1976, pp. 491–625.
7. Krüger, F. W. Metabolism of nitrosamines *in vivo*. II. On the methylation of nucleic acids by aliphatic di-n-alkyl-nitrosamines *in vivo*, caused by β-oxidation: the increased formation of 7-methylguanine after application of β-hydroxypropyl-propyi-nitrosamine compared to that after application of di-*n*-propyl-nitrosamine. Z. Krebsforsch. 79: 90–97 (1973).
8. Scribner, J. D., and Ford, G. P. *n*-Propyldiazonium ion alkylates O^6 of guanine with rearrangement, but alkylates N-7 without rearrangement. Cancer Letters 16: 51–56 (1982).
9. Krüger, F. W. Metabolismus von Nitrosaminen *in vivo*. I. Über die β-Oxidation aliphatischer Di-*n*alkylnitrosamine: die Bildung von 7-Methylguanine neben 7-Propyl bzw. 7-Butylguanin nach Applikation von Di-*n*-propyl- oder Di-*n*-butylnitrosamin. Z. Krebsforsch. 76: 145–154 (1971).
10. Leung, K.-H., Park, K. K., and Archer, M. C. Methylation of DNA by *N*-nitroso-2-oxopropylpropylamine: formation of O^6 and 7-methylguanine and studies on the methylation mechanism. Toxicol. Appl. Pharmacol. 53: 29–34 (1980).
11. Langenbach, R., Kuszynski, C., Gingell, R., Lawson, T., Nagel, D., Pour, P., and Nesnow, S. C. Structure-activity analysis of propyl and related nitrosamines in the hepatocyte-mediated V79 cell mutagenesis system. In: Structure-Activity Correlation as a Predictive Tool in Toxicology, Fundamentals, Methods and Applications (L. Golberg, Ed.), Hemisphere Press, New York, 1983, pp. 241–256.
12. Rose, S. L., and Jurs, P. C. Computer-assisted studies of structure-activity relationships of *N*-nitroso compounds using pattern recognition. J. Med. Chem. 25: 769–776 (1982).
13. Dunn, W. J., III, and Wold, S. An assessment of carcinogenicity of *N*-nitroso compounds by the SIMCA method of pattern recognition. J. Chem. Inf. Comput. Sci. 21: 8–13 (1981).
14. Dunn, W. J., III, and Wold, S. The carcinogenicity of *N*-nitroso compounds: a SIMCA pattern recognition study. Bioorg. Chem. 10: 29–45 (1981).
15. Jurs, P. C., Hasan, M. N., Henry, D. R., Stouch, T. R., and Whalen-Pederson, E. K. Computer-assisted studies of molecular structure and carcinogenic activity. Fund. Appl. Toxicol. 3: 343–349 (1983).
16. Kier, L. B., and Hall, L. H. Molecular Connectivity and Drug Research. Academic Press, New York, 1976.
17. del-Ré, G. A simple M.O.-L.C.A.O method for calculating charge distribution in saturated organic molecules. J. Chem. Soc. 1958: 4031–4032.

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.